

Survival and What Matters

Undergraduate Research Thesis

Presented in Partial Fulfillment of the Requirements for graduate “with Honors Research
Distinction in Philosophy” in the undergraduate colleges of The Ohio State University.

By

John Conlon

The Ohio State University

April 14, 2015

Project Advisor: Professor Donald Hubin, Department of Philosophy

I. Introduction

Survival matters a great deal to us. But what must be the case about a person in the future, for me to survive as that person? The intuitive answer is, simply, that this person be *me*. It seems, that is, that I survive as a person in the future if and only if that person and I are one and the same person. In still other words, it seems that what matters is identity. But Derek Parfit has convincingly argued that this cannot be right.¹ Suppose that I am one of the few people thought to exist that have perfectly symmetrical brains (in terms of abilities, memories, etc.). Now suppose that each half of my brain is transplanted into one of the bodies of my recently brain-dead, but otherwise unharmed identical siblings. Both wake up, and seem to remember my life, have my personality, etc. The question of personal identity would seem to have three possible answers. Either I am both of the resulting people, or I am only one of them, or I am neither of them. If I am identical to both, then, by the transitivity of identity, they are identical to each other. But they are clearly distinct. What's more, there seems to be no reason to say I am one and not the other; I stand in exactly the same relationship to each of them. We seem left with the conclusion that I am neither of them. But surely this event is not bad like ordinary death is bad. Identity, Parfit claims, is thus not what matters in survival. I, and many others, find this argument quite compelling, and I shall assume its conclusion henceforth.

I shall talk in this paper of death and survival. I shall use these terms in a way that does not presuppose identity. Thus one could survive, in my sense, even if there is no one in the future who is identical to oneself.

The question of survival is tightly linked to what some philosophers have called prudential concern. Prudential concern is the type of concern one has, in everyday life, about

¹ Parfit, Derek. *Reasons and persons*. Oxford University Press, 1984.

one's own future. One eagerly anticipates one's future joys and dreads one's future hardships. If my brain were divided and transplanted into my siblings' bodies, it seems that I could be concerned in this way about what happens to each, despite the fact that I am not identical with either. So identity is not what grounds prudential concern. I shall alternate between talk of survival and talk of prudential concern. As I use the terms, one survives as a certain person if and only if one has reason to be prudentially concerned about that person.

Throughout this paper, I shall be assuming what has been called Reductionism about personal identity; that is, I assume that there is nothing to my identity or survival through time over and above facts about my brain, my body, my mental states, and the relations between these things. So, for example, I shall assume without argument that there are no immaterial souls whose continued existence we must track to know facts about identity or survival. To decide questions of survival on a reductionist account, it will be sufficient to know facts about my brain, my body, my mental states, and the relationships between them. But to decide between reductionist accounts, we shall need to discover which of these facts matters to us in the way that survival does.

In this paper, I consider two of the most common types of reductionist accounts of survival and compare their relative costs. First, I discuss a range of views that fall under the category 'Psychological Accounts'. Roughly, these accounts define survival only in terms of mental states and the relations between them. I put forth several principles that, taken together, are incompatible with each of the Psychological Accounts of which I am aware. I find the denial of any of these principles absurd, and so I reject all of the Psychological Accounts. Others may not find them absurd, but in any case I hope to show the intuitive costs associated with such views. Next, I look at alternative, 'Physical Accounts', which assert that physical continuity of

the brain is a necessary condition for survival. I consider some arguments against such views, but I find them less than conclusive. Thus I conclude by endorsing the Physical Accounts over their Psychological rivals. I give two versions of Physical Accounts that seem plausible, but I do not know how to choose between them.

II. Psychological Accounts

To begin, consider

The Brain State Transfer (BST): A person, A, is about to enter a machine. Its owner has told her that it will operate in the following manner: it will scan and record the states of the brain of another person, B. Then, it will operate on her brain such that it is in the exact state that B's brain had previously been in. Thus, the person who wakes up with what had previously been A's brain and body, whom let's call the 'A-body-person', will have apparent memories qualitatively identical to B's memories. Similarly, the A-body-person's beliefs, desires, and other psychological characteristics will be exactly similar to how B's had been before the operation. B undergoes an analogous operation such that the person who wakes up with what had previously been B's brain has apparent memories, beliefs, character traits, desires, and so on, which are qualitatively identical to A's. Call this person the 'B-body-person'. The A-body-person thus will seem to remember entering the machine with B's body, and likewise the B-body-person with A's body. Each will seem to others overwhelming like the person who had entered the machine with the other body, except, of course, for the change in appearance. A believes the machine's owner that this is how the machine will operate.

Before this operation, the machine's owner tells A that he will torture one of the people who exits the machine and reward the other with a large sum of

money. But he tells A that she may choose who will be tortured and who will be paid. He says that he will carry out her decision, and she believes him.²

Suppose that A decides who will be tortured and who will be paid entirely on selfish grounds. What should she choose? How should she allocate the torture such that she need not dread it? To whom should she give the money such that she can look forward to spending it? In other words, which of the resulting people does she survive as?

Locke argued that personal identity through time essentially involved memory.³ That is, a person at one time is identical to a person at another time if and only if the later person remembers or could remember experiences had by the earlier person. Though we have decided that identity is not what matters, could a translated version of Locke's view provide a plausible criterion for survival? Does one survive if and only if there is someone in the future who remembers or could remember one's current experiences?⁴ Call this account of survival the Memory Account.

Reid gave the famous case of the Brave Officer to counter Locke's theory of identity.⁵ Though we are now considering the Memory Account of survival, not identity, Reid's objection is still troubling. Consider a brave officer who remembers a flogging he received as a youth, and an old general who remembers his brave deeds as an officer. We

² This example is a hybrid of Shoemaker's discussion in his and Swinburne's "Personal Identity" and of a similar example in Williams's "The Self and the Future."

³ In his "Of Identity and Diversity", which can be found in John Perry's collection *Personal Identity* (1975).

⁴ There is a concern here about when exactly one 'could remember' an experience. Locke clearly wants to allow for people to be asleep or to momentarily forget while still preserving identity. Since I eventually reject the Psychological Accounts anyway, I shall simply assume going forward that we have an adequate account of what Locke means here. But it is worth noting that for the advocates of such views, providing an analysis of "could remember" seems urgent.

⁵ In his "Of Mr Locke's Account of our Personal Identity", found in Perry (1975).

can all agree that the boy survives as the brave officer and the brave officer as the old general. But if we accept the Memory Account, we must say that the flogged boy does not survive as the old general. This is hard to accept.

We can revise the Memory Account to accommodate this intuition. We can now say that a person X survives as a person Z if and only if at least one of the following conditions holds:

- (1) Z remembers or could remember X's experiences.
- (2) X survives as a person Y, whose experiences Z remembers or could remember.

The Memory Account now does not require that there be someone in the future who remembers or could remember one's current experiences in order for one to survive. Rather, it requires that there be overlapping chains of such memories.⁶ Call the existence of such overlapping chains 'continuity' of memory. This new Memory Account accords somewhat better with our intuitions, but is it a plausible criterion for survival?

Some, like Butler, have argued that memory presupposes identity.⁷ If so, memory cannot be the criterion for survival; we have already seen, as in the introductory case of fission, that one could survive as another person without being identical to that person. Perry argues that memory does not presuppose identity, but rather some causal process

⁶ I shall leave out the "could remember" clause from here on out, since it makes for tiresome reading. Strictly speaking, this sentence should read something like, "Rather, it requires that there be a series of overlapping pairs of people who are such that the later person remembers or could remember the experiences of the earlier." Henceforth, I shall use "remembers", "memory", and so on to mean "remembers or could remember", "memory or the ability to remember", and so on.

⁷ In "Of Personal Identity", found in Perry (1975).

(presumably neurological).⁸ If this is right, then we could say that the two people who result from my fissioning would both remember my experiences. Their apparent memories and my earlier experiences are causally related in the right way.

If memory presupposes a certain causal process, it seems likely that cases such as the BST will exclude the possibility of memory. Presumably, brain scans and high-tech machines are not a part of the causal process that our concept of memory presupposes. And yet, many are drawn to the intuition that, after undergoing BST, A would survive as the B-body-person. The B-body-person does not, strictly speaking, remember A's earlier experiences, but many, such as Parfit and Shoemaker, seem to think that what she has is just as good as memory. Parfit introduces a new term, 'q-memory', to refer to the apparent memories that the post-BST people have. A person X q-remembers an experience of person Y if and only if

- (1) X seems to remember having an experience,
- (2) Y did have an experience like that X's apparent memory seems to represent.
- (3) X's apparent memory is causally dependent, in the right kind of way, on Y's experience.

Presumably, the right kind of causal dependency specified in (3) is weaker than that for memory, allowing that the B-body person could q-remember A's past experiences. Could q-memory be the criterion for survival? That is, does a person survive if and only if there is someone in the future who q-remembers her current experiences? Call this view the Q-memory Account.

⁸ "Personal Identity, Memory, and the Problem of Circularity", in his (1975).

Here, I also define the term 's-memory'. A person X s-remembers an experience of Y if and only if conditions (1) and (2) for q-memory hold. S-memories are just apparent memories of experiences. Thus, q-memories are a subset of s-memories. They are the s-memories that are caused, in the right sort of way, by the events they seem to represent. The Q-memory account thus states that a person survives as a future person if and only if there exists an overlapping chain of s-memories between her and the future person, and those s-memories are causally dependent, in the right sort of way, on the experiences they seem to represent. The reason for introducing this terminology will become clear later on.

The Q-memory Account gives what seems to many the correct answer to the BST case. But most philosophers do not find it a plausible criterion for survival. Many believe that, while q-memory does matter, it is not all that matters. Q-memories are just one type of what they call 'psychological connections'. The concept of a psychological connection is rarely explicitly defined; rather is illustrated by enumeration, and I shall follow that convention. The carrying out of an earlier intention and the persistence of memories, beliefs, desires, values, and character traits are paradigmatic examples of psychological connections. Roughly, if a person A is strongly psychologically connected to an earlier person B, then it will seem to A as though she had B's life. Like with memory, many believe that what matters is not just the holding of direct psychological connections; the holding of overlapping chains of psychological connections, or 'psychological continuity', also matters.

Like memory, saying that someone has carried out an earlier intention or that a belief or desire has persisted may presuppose some specific causal dependency of the later states on the earlier ones. And again the BST case may not satisfy the causal requirements for the other types of psychological connections. We might ask, "Did the B-body person carry out

the very same intention that A formed, or merely an exactly similar intention?" To avoid such concerns, it may be necessary to introduce terms such as 'q-intentions', 'q-desires', and so on, which have some less restrictive causal requirement. I also introduce here the corresponding terms 's-intentions', 's-desires', and so on, which have no requirement of causal dependency.

We now have the terminology to state Parfit's main thesis, that what matters in survival is Relation R. Relation R is psychological connectedness and/or continuity, with the right kind of cause. Parfit gives three candidates for the right kind of cause for Relation R, which are

- (1) its normal cause,
- (2) any reliable cause,
- (3) any cause.

Parfit notes that (2) does not seem plausible. Imagine an unreliable BST machine that only works one tenth of the time; the other nine out of ten times, it kills both people. Parfit claims, and I agree, that if the reliable BST machine allows for survival, then so does the unreliable one on the occasions where it in fact works. Parfit says of a similar case that, when it works, "this effect is just as good, even though its cause was unreliable". I shall ignore (2) for the remainder of this paper, as I and others believe it untenable. Parfit claims that we should accept (3), that what matters is Relation R with any cause.

Note however that Parfit defines psychological connectedness, and thus Relation R, in terms of q-memories and the persistence of, q-beliefs, q-intentions, etc. These concepts presuppose causal dependency of the later states on the earlier ones. I do not define psychological connectedness in this way. Instead I define psychological connections in

terms of s-memories and the persistence of s-beliefs, s-intentions, etc. I call psychological connectedness and/or continuity, in my sense, Relation S. I call Relation S with any sort of causal dependency Relation S + C1 and Relation S with the normal causal dependency Relation S + C2. Parfit's view can now be restated with this new terminology in mind; he believes that what matters is the holding of Relation S + C1.

Is this view tenable? Consider the following thought example, which is adapted from McMahan.⁹

Replication: Machines called Replicators exist that can, given a 'blueprint', create brains and bodies of any type, including with any set of psychological traits. There also exist machines called Scanners that can record the exact states of the cells of a brain and body. While it is prohibitively expensive to make a blueprint from the output of a Scanner, scientists have nonetheless developed a procedure for creating replicas of people. They have a program that, at virtually no cost, randomly generates blueprints, with random appearances and random sets of beliefs, desires, apparent memories, etc. They create very many of these random blueprints, and they assign an ID number to each. When someone wants to undergo replication, the scientists then try to find a blueprint that matches the output of the Scanner. If a match is found, the patient is given the ID number of the blueprint that matches the Scanner's output. She enters that number into the Replicator, her brain and body are destroyed, and then

⁹ McMahan, Jeff. *The ethics of killing: Problems at the margins of life*. Oxford University Press, 2002.

the Replicator creates, out of new matter, a brain and body according to the blueprint whose ID number she entered.

A and B each have a rare brain disease that, though they as yet experience no symptoms, will shortly kill them. The scientists have found blueprints for each that are exactly similar to them except that the resulting replicas will not have the brain disease. By an improbable coincidence, the ID numbers they are assigned are exactly the same except for the final two digits, which are transposed.

There are two scenarios I wish to consider:

- (1) *Normal Replication*: A and B each enter a Replicator and enter the numbers they were told. Therefore, replicas of each are created, which are exactly similar to A and B except that they do not have the brain disease.
- (2) *Accidental Replication*: B, by accidentally transposing the last two digits of the ID number he was told, enters the number that A was told. A, by a similar mishap, accidentally enters B's number. Therefore, replicas of each are created, which are exactly similar to A and B except that they do not have the brain disease.

If we believe that survival consists in the holding of Relation S + C1, we should say that A and B survive in the case of Normal Replication. A and her replica are strongly S related, and her replica's beliefs, memories, desires, and so on are causally dependent on those of A. A's having the beliefs, memories, desires, and so on that she did caused the scientists to tell her the ID number they did. Her entering that ID number into the Replicator caused her replica to exist. A corresponding story can be told about B and his replica. However, in the case of Accidental

Replication, there is no causal dependency between the mental states of A and those of her replica. Therefore, Relation S + C1 does not hold between A and her replica, and similarly for B and his replica. We must therefore say that what matters in survival is entirely absent. Normal Replication, on this account, is as good as ordinary survival, and Accidental Replication is as bad as ordinary death.

This is hard to accept. I claim, with McMahan, that if what matters is present in one case, it is present in the other. There is no difference between Normal and Accidental Replication in who enters or leaves the Replicator. And I cannot bring myself to believe that the difference between survival and death lies is the bizarre coincidence that distinguishes these cases. If we were friends with A and B, and discovered the odd coincidence that caused their replicas to exist, how should we respond to this news? If we believe that what matters is Relation S + C1, we should believe that it was a tragedy (or at least, a tragedy for A and B). “If only,” we ought to say, “they had entered the correct numbers! Then our friends would have survived, though the replicas who exist now would be no different. But now we must hold a funeral; be sure to invite their replicas.” This reaction seems to me absurd. To capture this intuition, I introduce the following principle:

Principle of Equivalence: What matters in survival is present in Normal Replication if and only if it is present in Accidental Replication.

This principle and the view that what matters is Relation S + C1 are incompatible. I find the denial of this Principle absurd. If this is right, Relation S + C1 is not what matters.

We can now move in either of two directions. We might say that what matters is present in neither case, or we might say that what matters is present in both. If we adopt the former position, we may say that there is some more restrictive requirement of causal dependency

that must obtain in order for one to survive. Alternatively, we could say that we were mistaken in thinking that there was any requirement of causal dependency at all; that is, we might say that what matters is just Relation S.

Though I am not certain, I believe that Parfit might accept this latter claim. I quoted him above as saying, in regard to a case where Relation R held due to an unreliable cause, that “this effect is just as good, even though its cause was unreliable”. It may be that he could accept that the outcome in Accidental Replication is just as good, though there is no causal dependency.

Before turning to the view that Relation S is what matters, we should note that Relation S can hold to different degrees. Memories fade and new ones replace them, intentions change, true beliefs (we may hope) replace false ones, and so on. I am very psychologically connected to myself yesterday, but I am perhaps only weakly psychologically connected to myself when I was six. Thus, psychological connectedness comes in degrees. Parfit defines psychological continuity as an all-or-nothing relation, but as McMahan notes, it need not be. We can say that one person is strongly psychologically continuous with another person if there exists a chain of strong psychological connectedness between them. Similarly, if there is only a chain of weak psychological connectedness between two people, we may say that the one person is only weakly psychologically continuous with the other.

Parfit believes that the rational level of prudential concern can vary with the degree to which Relation S holds.¹⁰ He writes that, “It can be rational to care less, when one of the grounds for caring will hold to a lesser degree.” This principle seems right to me, and I shall assume it going forward.

¹⁰ Of course, he actually phrases it in terms of Relation R. But we have abandoned that terminology.

Let's turn now to the thesis that what matters in survival is the holding of Relation S. McMahan argues that this view has the absurd result that we ought to be prudentially concerned about other people in our own society in the actual world. His reasoning is as follows. We share a great many beliefs with most other people in our society. For example, I and very many other people all believe that the capital of Ohio is Columbus, that Venus is the second planet from the sun, that plants photosynthesize whereas animals do not, and countless other facts. I and the same people may share the desires to one day compete on *Jeopardy!*, to order pizza for dinner sometime this week, and many other desires. Thus, it seems we share many s-beliefs and s-desires; presumably similar examples could be filled in for other types of psychological connections. Having recognized this, aren't we committed to saying that I have reason to be prudentially concerned about these people's futures? We may admit that my level of prudential concern for them may be significantly less than that which I feel for my own future, seeing as I shall be significantly more psychologically connected with myself in the future than with them. But are we prepared to say that I have any reason at all to be prudentially concerned for these other people? This seems absurd. If this line of reasoning is right, then Relation S is not what matters.

I believe that the view that Relation S is what matters can withstand McMahan's argument. We have accepted that Relation S comes in degrees and that the appropriate level of prudential concern can vary with the degree to which Relation S holds. But we need not say that the appropriate level of prudential concern is nonzero whenever the degree to which Relation S holds is nonzero. We might say that there is a threshold, below which there is no reason for prudential concern, and above which the appropriate level of prudential concern is positive and

varies with the degree to which Relation S holds.¹¹ We might claim that the degree to which Relation S holds between any two people in the actual world is well below this threshold and that, therefore, there is no reason in the actual world to be prudentially concerned about other people.

Here is an analogous case. I have fond memories of a vacation my family once took to Maine. I remember looking out onto the Atlantic Ocean from a particular patch of rocky shoreline as the waves sallied and retreated. Imagine that, because I enjoy thinking about this memory, a friend of mine decides to buy me a painting that she hopes will remind me of that place. Surely, it would be best if she could find a painting that looks exactly similar to that stretch of the Acadian coast. It would also be good, though less so, if she could find a painting that looked somewhat similar. Thus it seems that the value of this painting might vary with the extent to which it resembles that particular shore I once visited. But would a painting of fish tank be valuable, since it depicts water? Would a painting of a fish tank that had some pebbles at the bottom be more valuable? That painting would include water and stones, so it seems that it would be more similar than the former picture, whose only claim to similarity was that it depicted water. Must we now also say that it would be better for her to buy this painting than the other, in virtue of this increased similarity? Surely not. We may admit that the value of the painting varies with how similar its scene is to the Acadian coast. And we may admit that the

¹¹ Here, I am using “appropriate level of concern” to mean something like “the level of concern one’s desires and interests give one derivative reason to have. For example, a baseball fan’s appropriate level of concern about the latest scouting reports may be zero, if those scouting reports were of football players. It isn’t that it would be irrational for her to care about football recruiting (though this would be a bizarre passion to have without a corresponding love of the game itself); rather, her interest in baseball gives her no derivative reason to pay attention to the goings-on in this other sport. Similarly, I do not mean to suggest that it would be irrational to have different preferences concerning survival. I merely mean that, given what we are supposing these preferences to be, this or that is the appropriate level of derivative concern.

painting of the fish tank with pebbles is more similar to that coast than the painting without pebbles. But we can also say that neither painting is sufficiently similar to be valuable. There may be a threshold of similarity, above which the painting's value increases with its similarity, but below which the painting is not at all valuable.¹²

I think the view that Relation S is what matters can survive McMahan's argument, but I believe that an argument similar to McMahan's is more difficult for it to withstand. Consider

Our Possible Multiplicity: The universe may be infinitely large. From what we can tell, on a wide enough scale, matter is homogeneously distributed throughout it. In every direction, there are millions of galaxies stretching as far as our instruments can detect. If the universe is infinitely large, and if galaxies populate space everywhere with a similar density with which they populate the observable universe, then it seems there must be infinitely many stars and planets. As evidenced by our own existence, the probability of a planet like ours existing is nonzero. So, it would seem that, if the universe is infinite, there must be other planets very similar to ours. Indeed, there must be other planets exactly similar to ours, and there must be infinitely many of them. So, if the universe is infinite, there are (were, and will be) people whose lives are phenomenologically indistinguishable from our own. We are fully S-related to those people.

If the universe is infinite, then there will be people in distant galaxies fully S-related to us. The universe may be infinite, and so there may be such people. If we believe that Relation S is what matters in survival, that is, what grounds prudential concern, then we should think that

¹² This analogy may go one level deeper. It may be that the painting becomes valuable only when its scene is somehow distinctive of the place I visited. After all, very many places have water, and very many places have water and stones. But perhaps the painting starts to become valuable when the list of places that it might be depicting grows sufficiently small. Similarly, perhaps the level of psychological connectedness only begins to matter for prudential concern when the psychological connections between a person at one time and another person at a later time are somehow distinctive of the earlier person.

we survive as those people. We should look forward to their future joys, and we should dread their future hardships.

I cannot bring myself to believe this. If the universe is infinite, then presumably there are many people fully S-related to me. More disturbing is the fact that there are presumably many ways the “local” universe could be such that someone fully S-related to me would exist.¹³

Consider

The Meteor Strike: Scientists announce two discoveries on the same day. First, they announce that it has been proven that the universe is infinitely large and that there are infinitely many planets, stars, etc. Second, they announce that they have discovered a meteor that will certainly strike the Earth tomorrow, destroying all life on the planet. However, they reason that, since the universe is infinitely large, all possible local universes are actual. Surely, they argue, it is possible to have a local universe exactly like ours except that a meteor corresponding to the one on a collision course with Earth does not exist. So, they conclude, there is a planet full of people completely S-related to us now who will survive the next day.

If we accept that what matters is Relation S, how should we react to these scientists’ news (assuming we believe them)? First, it seems we should regard the meteor’s impending strike with indifference.¹⁴ Even though the meteor will destroy the Earth, there will still be people tomorrow fully S-related to us. And since that is simply what we mean by survival, we should say that we shall survive the destruction of the Earth.¹⁵

¹³ By ‘local universe’, I simply mean a large enough expanse of space. For example, the sphere 14 billion light-years in radius with the Earth as its center (roughly, the observable universe) is a local universe. If the universe is infinitely large, there will be infinitely many equally large, non-overlapping spheres.

¹⁴ Putting aside concerns other than survival, of course, such as the suffering of those on Earth. We can assume that the meteor kills everyone on Earth almost instantly.

¹⁵ One might reply, “But there will be fewer such people, so we may in fact care about the destruction of the Earth.” Perhaps. But, in this thought experiment, there are very many (indeed,

It is less clear to me how relieved we ought to feel regarding the news that the universe is infinitely large. We shall survive the meteor strike, but this is not the whole story. There is a planet on which things go on much as they had been on Earth except that no meteor hits, but that is not the only other local universe containing people S-related to ourselves. Presumably, there are many, many more. It would seem that we should look forward to experiencing each of them. In some of these local universes, a smaller meteor hits and causes great suffering but fails to render life on Earth extinct. In others, no meteor hits, and the person S-related to me wins the lottery. In yet others, the person S-related to me contracts a lethal cancer that shortly kills him. There are countless more possibilities. Could I anticipate surviving as all these people? In the case of fission that begins this paper, there are two people whom I shall survive as. Though it is counterintuitive, I can accept that I could look forward to surviving as each of them. But if what matters is Relation S, then, if the universe is infinitely large, I should look forward to surviving as countless millions of people, many of whom may have vastly different futures and all of whom will live trillions of miles from here and from each other. I cannot believe this.

I introduce the following principle, whose denial seems to me absurd, to rid us of these concerns.

Principle of Locality: The objects and events that are outside the range of causal interaction with the Earth have no bearing on our survival, no matter what is out there.

infinitely many) people whom we shall survive as. Could the loss of just one of these people (namely, the one here on Earth) matter all that much? Surely, at least, it could not matter as much as we often consider our deaths to matter. So, perhaps I should say that we should regard the destruction of the Earth not with indifference, but with something very close to indifference.

This principle and the view that Relation S is what matters are incompatible. Because I am forced to choose between them, I choose the Principle of Locality and reject the view that Relation S is what matters.

Earlier, I argued that, if we accept the Principle of Equivalence, we could move in one of two directions. The first was to claim that what matters is not Relation S + C1, but rather just Relation S. I have argued that, unless we are willing to reject the Principle of Locality, this view is untenable. The second path forward was to accept some stricter causal requirement between a person at one time and an S-related person at another time for the former person to survive as the latter. I shall start with the view that what matters in survival is Relation S with its normal cause: that is, Relation S + C2. I do this simply because, to my knowledge, it is the only view that incorporates Relation S and some stricter causal requirement.¹⁶

Physical continuity of the brain is a necessary condition for Relation S + C2, but it is not sufficient. Parfit seems to ignore this fact. For example, he writes, “It might be suggested that what matters is *both* Relation R *and* physical continuity. But this is the same as Relation R with its normal cause, since physical continuity is part of R’s normal cause” (emphasis his). Later he writes of a fissioning case that, “[his] relation to each of the resulting people is Relation R with its normal cause, enough physical continuity.”

This view, that Relation S + C2 and Relation S plus physical continuity are the same, is false, as the following thought experiment shows. Consider

The Reprogrammer: A mad scientist, no doubt interested in the puzzles of personal identity, has developed a machine, which works in three stages. First, it records the states of a person’s brain exactly. Then, using super-high-tech-neuro-

¹⁶ The exception being Parfit’s brief discussion of Relation R with a reliable cause. I ignore this view because Parfit rejects it quickly and, I believe, with good reason. I know of no authors who assert it.

technology, it changes that person's brain such that the resulting person is no longer S-related to the original person.¹⁷ Finally, using the data it received from recording the person's brain, it changes the person's brain back to its original state. The resulting person is completely S-related to and physically continuous with the original person. Further, this 'Reprogrammer' can perform this procedure on people remotely, and this operation lasts less than one second.

There is physical continuity of the brain between the pre- and post-operation people. They are also fully psychologically connected, but this is not due to its normal cause.

I believe it is conceivable (though whether it is in fact possible is another question) that one could undergo Reprogramming while conscious and without noticing it. I believe this because it seems to me that the vast majority of our mental states often do not 'feel like anything'. For example, there is nothing that it is like to remember my family's vacation to Maine when I am not thinking of it. There is nothing it is like to love the ones I love when I am not thinking of them and they are not present. And so on with almost all my psychological traits. Thus, I cannot claim that if, for a fraction of a second, all these mental states were to change, only to be quickly replaced by exactly similar ones, I would notice this happening. If this is right, let us imagine that the Reprogrammer operates upon people while they are awake and that people do not notice its operation.

Relation S holds between the pre- and post-Reprogrammer people, but this relation does not hold due to its normal cause. If we believe that what matters in survival is Relation S + C2, we must therefore believe that someone about to undergo Reprogramming should view this

¹⁷ If we believe, as discussed above, that there may be a threshold above which S-relatedness begins to matter, then the machine need only change the person's brain such that the resulting person is S-related to the original person, but only to an extent just below that threshold.

prospect as just as bad as ordinary death. This seems absurd. Again, I introduce a principle to handle this intuition.

The One-Second Principle: If there is a person one second from now who has what is unambiguously my brain and my body and who is completely S-related to me, then I have survived.

This Principle and the view that what matters is Relation S + C2 are incompatible. Since I am forced to choose between them, I choose to accept the One-Second Principle and reject the view that what matters is Relation S + C2.

I began by discussing Relation S + C2, but it may be that there is some other causal requirement for survival, more restrictive than C1 but less so than C2. Call this requirement, whatever it may be, C*.¹⁸ The One-Second principle is also incompatible with the view that what matters is Relation S + C*. Consider

The Random Reprogrammer: The Random Reprogrammer works just as the original Reprogrammer, with one exception. Instead of changing the person's brain back to the state it was in previously, it changes the brain such that it has a randomly generated set of q-memories, q-beliefs, q-intentions, and so on. The mad scientist uses the Random Reprogrammer on me. By a bizarre coincidence, the randomly generated mental states the machine selects are exactly similar to those I began with. Again, this operation happens remotely and takes less than one second. I do not notice its occurrence.

In this case, there is someone one second from the beginning of the operation who is completely S-related to me and who has what is unambiguously my brain and my body. But, there is no causal dependency between my mental states prior to undergoing Random

¹⁸ Note that, because C* is arbitrary, I cannot assume that the original Reprogrammer does not fulfill C*.

Reprogramming and the mental states of the post-operation person. Thus, Relation $S + C^*$ does not hold. So, if we believe that what matters is Relation $S + C^*$, we must say that I do not survive the Random Reprogrammer. This assertion violates the One-Second Principle. Since I must choose between them, I choose to accept the One-Second Principle and reject the view that what matters is Relation $S + C^*$.¹⁹

I have given four versions of the Psychological Account: Relation S with no causal requirement and Relation S plus $C1$, $C2$, and C^* . All of these are incompatible with at least one of the three principles I have put forth. There may be other formulations of the Psychological Account other than those I've given that avoid these issues. But, to my knowledge, the accounts I've given exhaust those in the literature. As mentioned above, Parfit favors the view that what matters is Relation R with any cause, or, in my terminology, $S + C1$. Lewis shares this view, as did Shoemaker, though he has since changed his mind.^{20,21,22} Shoemaker now favors the view that what matters is Relation $S + C2$, as does

¹⁹ Actually, the One-Second Principle is also incompatible with the view that what matters is Relation $S + C1$. However, though I find the One-Second Principle compelling, I find the Equivalence Principle even more so. Thus, I'd like to rely on the latter as much as possible, which is why it remains in this paper.

²⁰ Shoemaker, Sydney, and Richard Swinburne. *Personal Identity*. Oxford, England: B. Blackwell, 1984. Print.

²¹ Lewis, David. "Survival and identity." (1976).

²² Shoemaker actually describes his notion of survival as the playing out of the causal roles of mental states over time. I do not know if this should be counted as saying that Relation $S + C1$ is what matters (though he believed that one could survive the BST). If this is an inappropriate classification, then, since Shoemaker's is certainly a causal account, it is incompatible with the One-Second Principle, as are all views that incorporate Relation $S + C^*$.

Agar.^{23,24} To my knowledge, no one has put forth a different causal account than these, and no one has put forth the view that what matters is Relation S alone.

III. Physical Accounts

I now turn to views that incorporate the physical continuity of the brain, what I shall call 'Relation P', as a necessary condition for survival. I shall consider the views that what matters is Relation P + S and that what matters is Relation P alone. I shall ignore Physical Accounts that incorporate C1, C2, or C*, as these views will still be incompatible with the One-Second Principle.

These two views, concerning Relation P + S and Relation P alone, are not incompatible with any of the three Principles that I have introduced. If Relation P is a necessary condition for survival, we should say that I survive neither Normal nor Accidental Replication. Then, we can accept the Equivalence Principle. Next, any objects or events that interact causally with my brain, and thus that affect the extent to which Relation P holds, are within the range of causal interaction with the Earth. So we can accept the Principle of Locality. Finally, these views are obviously compatible with the One Second Principle, since having the same brain one second from now entails that Relation P holds.

²³ Shoemaker, Sydney. "Functionalism and personal identity—A reply." *Noûs* 38.3 (2004): 525-533.

²⁴ Agar, Nicholas. "Functionalism and personal identity." *Nous* 37.1 (2003): 52-70. Agar believes that one person A survives as another person B if between their mental states there is a chain of causes with no non-psychological intermediates. Shoemaker agrees with this view. This strikes me as something similar to Relation S with its normal cause, but even if it is not, this view is nonetheless incompatible with the One-Second Principle.

I find the Physical Accounts intuitively quite attractive, and I am not alone. McMahan discusses a thought experiment he calls Double Replication. This case is exactly like the Replication case that I have described, except that the Replicator makes two duplicates of a person, instead of just one. He compares this case to Parfit's fission case, which introduces this paper. He writes,

Yet, had Parfit presented the phenomenon of division with [Double Replication], he would have been unable to persuade many of his readers to accept his conclusion. ... [M]ost of us agree that what matters is present in the relation between the original person and both his successors in the case of Division; but many of us do not find a basis for [prudential] concern in the relation between the original and his replicas in the case of Double Replication (pg 59).

Even those who defend a Psychological Account admit to harboring Physical sentiments. Shoemaker, who believed that what matters is Relation S + C1, writes,

If I am asked why I regard the continued normal functioning of the brain as a better realization of psychological unity than the operation of the BST-device, I can give no reason. And I can give no reason why I would prefer (as I now can't help feeling that I would) to survive in the ordinary way, with my present brain and body, rather than to have my brain-states transferred to a healthy clone of my body". (pg 132)

Note that Shoemaker seems to be concerned here about whether we can give a reason for our preferences concerning our survival. I am less concerned about this issue. I do not think that it is irrational to hold preferences without reason. I shall not argue much for this position; instead I shall merely note that I can think of no reason why I prefer knowledge to

delusion, pleasure to pain, love to apathy.²⁵ I simply do. But I do not think this lack of an ability to give a reason renders these preferences irrational.

Why, then, if there does seem to be this deep-seated intuition about physical continuity, do so many philosophers reject the Physical Accounts? Parfit, though he sometimes admits Physical intuitions, believes that “physical continuity is the least important element in a person’s continued existence.” He writes,

Why should my brain be singled out in this way? The answer must be: ‘because the brain is the carrier of psychological continuity, or Relation R’. If this is why the brain is singled out, the continuity of the brain would not matter when it was *not* the carrier of Relation R. The continuity of the brain would here be no more important than the continuity of any other part of the body. And the continuity of these other parts does not matter at all.

Parfit is assuming here that the only reason we might single out the survival of the brain as a candidate for what matters in survival is that it is the carrier of Relation R. This claim is, I believe, false. Consider the following argument.

I am a thinking thing. My brain is a thinking thing. There is only one thinking thing present in this room. I and my brain are both in this room, so I am my brain. Though identity is not what matters in survival, the relation underlying identity is what matters. Because I am my brain, the relation underlying my identity through time is the relation that

²⁵ I am, of course, ignoring instrumental reasons for preferring these things. For example, one reason I prefer knowledge to delusion is that having true beliefs better allows me to pursue my other desires. No doubt this reason at least partly explains, from an evolutionary standpoint, why I prefer knowledge. I am claiming, instead, that I can give no reason why I should *intrinsically* prefer knowledge to delusion.

underlies my brain's identity through time. This relation is physical continuity. So, what matters in my survival is physical continuity of this brain.

I do not claim that this argument is sound; I claim only that it is intuitively plausible and makes no mention of Relation R, S, C1, and so on. Therefore, unless decisive counterarguments are forthcoming, we should deny Parfit's premise that the only reason the brain might be singled out is that it is the carrier of Relation R. Thus, we can reject his argument.

But this is not Parfit's only argument against the view that what matters is Relation P. He asks us to consider

The Operation: I have a brain defect that, if untreated, will kill me. The doctors have told me that they would be able to replace every part of my brain with a replica that is exactly similar except that it lacks this defect. They may do so in one of these two ways:

Case One: My entire brain is removed and replaced all at once with an exactly similar brain.

Case Two: First, only one one-hundredth of my brain is removed and replaced by an exactly similar part. Then, after some length of time t has passed, another part is removed and replaced. This process is repeated until all the parts of my brain have been replaced with exactly similar parts.

Parfit claims that Case Two preserves physical continuity, while Case One does not. If this is right, then, if we believe that Relation P is a necessary condition for survival, we should claim that one could survive Case Two but not Case One. But, Parfit claims, the difference between survival and death cannot consist in the difference between these two

cases. He asks, “Can *my* fate depend on this difference in the ordering of removals and insertions? Can it be so important, for my survival, whether the new parts are, for a time, joined to the old parts?” Let us introduce a principle to accommodate Parfit’s intuition:

Parfitian Principle: If every part of my brain is to be replaced with an exact replica, it cannot matter for my survival whether the new parts are joined for a time with the old parts.

If Parfit’s argument so far is correct, then the Parfitian Principle is inconsistent with the view that what matters is Relation P. Parfit chooses to accept the Principle and therefore to reject the view that what matters is Relation P.

McMahan simply denies the Parfitian Principle. He writes, “I suspect that most people’s intuitions are different [from Parfit’s]... In short, to many of us, myself included, it does not seem arbitrary to suppose that the difference between Case One and Case Two is crucial to whether there is a basis for [prudential] concern.” McMahan then considers a spectrum of cases between Case One and Case Two, wherein the value of t , the time between operations, varies. I quote him at some length:

If, for example, fifty replacements were carried out each day, over a period of two days, it seems highly doubtful ... that what matters would be preserved. If however, the replacements were very gradual ... it seems compelling both that there would be continuity of the relevant parts of [the] brain and that [one’s] grounds for [prudential] concern about the future would, in each case, be preserved. There would be physical continuity, despite the replacements, because our understanding of the continued existence of a physical object normally tolerates a replacement of the object’s constituent elements or parts, provided no one of the parts is essential to the existence of the whole and that the turnover is sufficiently gradual that each new tissue would coexist for a

significant period of time with substantially greater amounts of older matter.
(pg 71)

McMahan therefore believes that it can in fact matter to my survival, if every part of my brain is to be replaced with exactly similar parts, whether the new parts are joined for a time with the old. He believes that the extent to which what matters is present is a function of t . As t increases, what matters is increasingly present in the relationship between myself before the operations and the resulting person after all of them.

I find McMahan's response compelling, though it is less an argument than a statement of intuitions. Parfit finds the denial of his Principle absurd; McMahan and I do not. We have our line in the sand.

In this paper, I have attempted to put pressure on the advocates of the Psychological Accounts. The three principles I introduced are, when taken together, inconsistent with each of the Psychological Accounts of which I am aware. I find the denial of any of these principles much more implausible than the denial of the Parfitian Principle. I therefore favor the Physical Accounts over the Psychological. I do not know how implausible defenders of the Psychological Accounts will find the denial of my principles, but in any case I believe I have bolstered the case in favor of the Physical Accounts.

Suppose we have agreed that physical continuity of the brain is necessary for survival. We still must decide whether it is sufficient. I believe that the only plausible additional necessary condition is Relation S. If this is right, then we must decide whether Relation P or Relation P + S is what matters.

I find it difficult to decide between these two views. It might help to look at cases where the two diverge. We could look at the BST case. We have decided that A does not

survive as the B-body person after the operation. But does she survive as the A-body person? Or has she died? In less drastic, though realistic cases, people with severe dementia or late-stage Alzheimer's disease are often not very psychologically connected with themselves as they were before they fell ill. Do they survive or not? I am unsure.

McMahan, when considering cases such as these, takes a middle road. He believes that Relation P is a necessary and sufficient condition for survival. However, he believes that Relation P alone is only sufficient for some small amount of prudential concern. Relation S, though neither necessary nor sufficient for survival, increases the appropriate level of prudential concern when conjoined with Relation P. Thus he answers that A, in the case of the BST device, has some small reason to be prudentially concerned about the A-body person, but none to be prudentially concerned about the B-body person. Similarly, he believes that, if I learned that I would someday become very mentally ill or develop Alzheimer's disease, I would have some small reason to be prudentially concerned about my future self in that condition. But this level of prudential concern would be much less than that which I should feel toward my future if I were not to develop such an illness.

IV. Conclusion

Perhaps McMahan is correct. But I cannot find any firm convictions with which to choose between these two views. Imagine that I was to undergo a Brain State Transfer, as discussed above. I have argued that I would not survive as the person to whom my brain-states are transferred. But between myself now and the person who will wake up from the procedure with my brain, Relation P will hold while Relation S will not. Shall I survive?

Should I regard this operation with fear, as my impending death, or as simply a drastic change in my character (though, of course, I could be dismayed about that as well)? It seems, as I consider it now, that the operation would carry great risk, but this is not quite right. I know exactly what will happen; someone will wake up with my brain and body, but without my memories, beliefs, desires, and so on. The only uncertainty is whether my relationship to that person contains what matters. Williams, in considering a similar uncertainty, found it disturbing.²⁶ I do as well.

²⁶ In his “The Self and the Future”, in Perry (1975).